Machine Learning Week 8

Unsupervised Learning

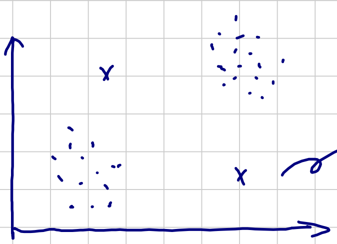
1. Clustering

group. / clustering:
find <u>some</u> structures

Training set $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$

Applications
- market segmentation
- social network analysis
- organize computer clusters
- astronomical data analysis

2. K-means Algorithm.



cluster centroids. (randomly initiated)

step (1) Cluster assignment.
- binary assignment. of datasets. depending on proximity.

(2) Move centroid.
- move to "mean" of location in all labelled

re-colour.

[Input].

① k (number of clusters)

② Training set $\{x^{(1)}, x^{(2)}, \dots, x^{(m)}\}$

$x^{(i)} \in \mathbb{R}^n$ [drop $x_0 = 1$ convention]

[Algorithm]

Randomly initialize K cluster centroids μ_k, $k = 1 \cdots K$

do {

for $i = 1 \cdots m$.

cluster assignment $\quad c^i := \min_k \| x^{(i)} - \mu_k \|^2$

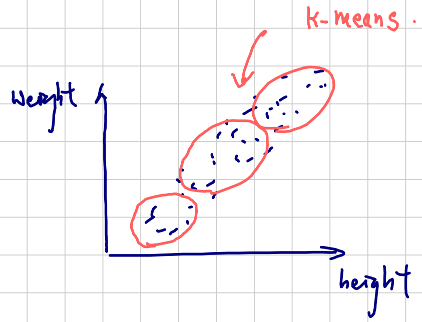
for $k = 1 \cdots K$

move centroid $\quad \mu_k := \text{mean}$ (pts assigned to cluster k)

$\in \mathbb{R}^n$

3. K-means for non-separated clusters.

T-shirt sizing



Optimization objective

K-means optimization objective.

- c^{i} : index.
- μ_k : cluster centroid k $\in \mathbb{R}^n$
- μ_{c^i} = cluster centroid of cluster to which $x^{(i)}$ has been assigned.

$$J(c^{(1)}, \dots, c^{(m)}, \mu_1 \cdots \mu_k) = \frac{1}{m} \sum_{i=1}^{m} \| x^{(i)} - \mu_{c^{(i)}} \|^2$$

$$\min_{\substack{c^{(1)} \dots c^{(m)} \\ \mu_1 \dots \mu_k}} J(\dots)$$

min.

$x^{(i)}$ x μ.

[Algorithm]

Randomly initialize K cluster centroids μ_k, $k = 1 \cdots K$

do {

for $i = 1 \cdots m$.

cluster assignment: $c^i := \min_k \| x^{(i)} - \mu_k \|^2$

Minimize $J(\cdots)$ wrt. $c^{(1)} \cdots c^{(m)}$. [holding $\mu_1 \cdots \mu_k$]

for $k = 1 \cdots K$

move centroid: $\mu_k :=$ mean (pts assigned to cluster k) $\in \mathbb{R}^n$

Min. w.r.t. $\mu_1 \cdots \mu_k$

Random Initialization

Rules.

(1) $K < m$

(2) Randomly pick K training examples

(3) set μ_k = examples

- Might have different clustering (local optimum) ⟶ try different random initialization.

- Implementation.

for $i = 1 \cdots 1000$ {

randomly initialize K-means
run K-means. get $c^{(1)} \cdots c^{(m)}$, $\mu_1 \cdots \mu_k$

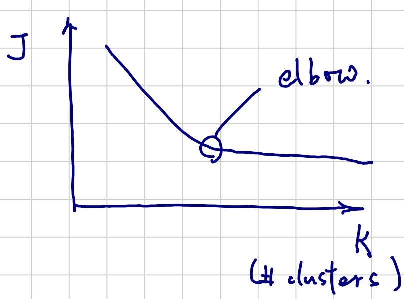
complete cost function (distortion)
↳ $J(c^{(1)}, \dots, c^{(m)}, \mu_1 \cdots \mu_k)$

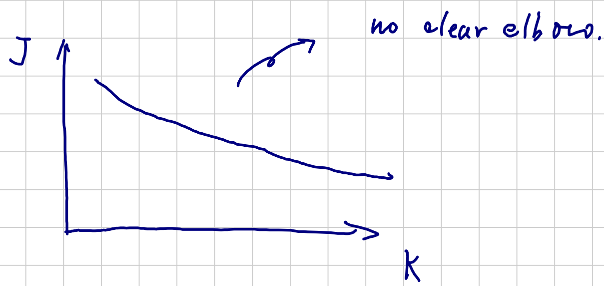
}

Pick one with lowest cost J.

Choose K

1. Elbow method:





2. Pre-defined (t-shirt sizes: 3 ... 5).